LAB I EXPERIMENTATION REPORT ON THE DETECTRON

Prepared By: Moshood Olawale Yekini
African Masters in Machine Intelligence, AMMI Ghana | *myekini@aimsammi.org*

1. INTRODUCTION

Recent advances in the field of computer vision courtesy of the use of deep neural network have engineered growth in many real-world applications in robotics, healthcare, agriculture, finance etc. This is as a result of research efforts in designing architectures, algorithms, softwares, frameworks and hardware that ease the efficiency of set of tasks in areas of interest in computer vision – classification, detection, segmentation and others.

Detectron is a PyTorch-based software system that uses the state-of-the-art object detection algorithms powered by Caffe2 deep learning framework. It robustness has enabled the experimentation of human-object detection, recognition, pose estimation and segmentation.

2. PLAYGROUND EXPLORATION

As default on the playground notebook, it loads a [Mask R-CNN][maskrcnn] model with a ResNet50 backbone pre-trained on the COCO dataset – a collection of annotated images with bounding boxes and 80 instance masked object categories, for object detection and instance segmentation. Widely enough, the predictions on these images for the task of instance segmentation and pose estimation achieved high level accuracy score on the custom ballon dataset.

3. EXPERIMENTAL REPORT AND OBSERVATION

We experimented the performance of the detectron on our collections of variant images (orientation, post styles, symmetric and others). The followings are reported observations on the two given tasks with some used sample images in which there are objections.

3.1 Instance Segmentation

Fig 1: Accurate instance segmentation of pregnant women holding booklets

Fig 2: Accurate instance segmentation of ladies performing yoga with inverted body orientation

Fig 3: Incorrect segmentation and prediction of handset as remote and handbag as banana

Fig 4: Incorrect segmentation and prediction of attire extension as handbag and segmentation of yam tubers with person

Model Architecture: We experimented Mask R-CNN with Feature Pyramid Network (it uses the ResNet50 backbone network that extracts feature maps from the input image at different scales on the COCO dataset) as bounding box detector because of its high sensitivity in detecting size ranging objects.

Examples and Observations: A very high accuracy is achieved in the explored images in the detection and instance segmentation of the inputted images which marks our choice of the architecture and the robustness of the detectron. However, there are cases of failure in this given task which are noted in the caption of the figures above.

Likely due to the pre-trained dataset which it was trained on, the model seems to be biased towards the dressed styled pattern (wedding pictures) …….. More so, there is a camera-focus issue as mostly close objects are assigned high accuracy score compare to far objects in the same input. Likewise, blurriness and noisiness had effects on the behavior of the model.

3.2 Pose Estimation

Model Architecture: We experimented keypoint with Feature Pyramid Network which uses the ResNet50 backbone also pre-trained on the COCO dataset in estimating poses in the inputted images.

Examples and Observations: The pose estimation is very accurate for detected human posture in the inputted images. However, there are noted instances when the pose fails in terms of its connectivity especially for connected images and cases when the pose is not estimated for some images.

Fig 3: Inaccurate pose estimation as it is connected to the bounding box instead of the rest of her body

Fig 1: Accurate pose estimation despite camera-focus proximity effect for all objects

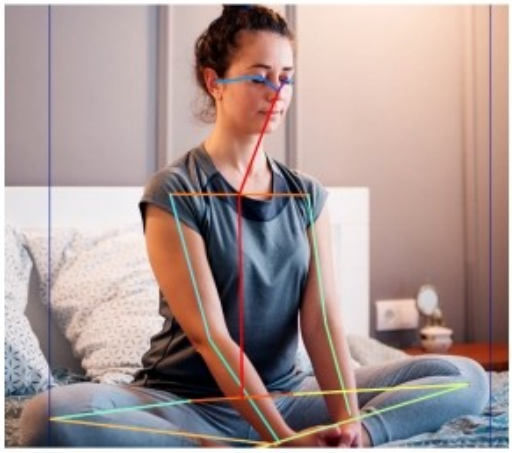

Fig 2: Accurate pose estimation with appropriate connectivity pattern of keypoints

Fig 4: Incomplete pose estimation and no pose detection for rear and side captured objects in the image

Error modes observed in the instance segmentation task also had effect the pose estimation also.

4. CONCLUSION

We experimented the Facebook AI Research (FAIR) software system – Detectron used for objection detection, instance segmentation and pose estimation. Our observation of its high prediction score and accuracy using our model's choice shows its robustness although failed in some adversarial examples – orientation, camera capturing positioning, scales, object similarities among others.

5. REFERENCES

- Georgia Gkioxari, Advanced Course in Computer Vision, AMMI Rwanda/Ghana, 2020
- Justin Johnson, Deep Learning for Computer Vision | Object Detection and Segmentation, University of Michigan, USA, 2019
- https://github.com/facebookresearch/Detectron